Synopsis of

"Usage of Statistical Applications and Its Secular Change in Bio-Medical Research"

Submitted for the award of degree

Of

Doctor of Philosophy

(In Subject of Statistics faculty of Science)

By

Swati. B. Patel

Under the guidance of

Dr.Vipin. D. Naik

M.Sc.(Statistics),M.phil , PhD.

Principal of S.P.B English Medium Collage of Commerce, Surat.

Submitted to

Veer Narmad South Gujarat University

Udhna- Magdalla Road,

Surat-395007, Gujarat (India)

Introduction:-

Statistics forms the basis for research as in virtually all research studies the whole universe is not covered and only a sample is covered. At the same time we need to know issues such as probability of our results holding true as compared to that obtained for the whole universe. Statistical methods play vital roles in the scientific research process. Over the past decades, a great increase in the use of statistics has been documented, for a wide range of medical journals (Altman, 1982, 1991, 2000). Thus, statistical analysis is an inseparable part of scientific publications, including the field of bio-medical research, biology and genetics. The results of researches have to be analyzed statistically in an adequate manner if the samples and/or measurements are obtained on more than one object, and/or more than once. Nowadays, this requirement applies to most of the scientific investigations as it helps to arrive closer to the core of the problem. Results of adequate statistical analysis though inspire new hypotheses but on the other hand, the results of an incorrect one are likely to cause errors and confusion.

Many reviews have been conducted in different countries, especially in Western countries to describe the use of statistical techniques in medical journals. Few reviews are also organized for the medical journals of developing countries. Almost all the reviews showed that there was a positive growth for the use of statistical techniques in medical writings. Better and more complex statistical analyses are introduced to prove the hypothesis.

Biomedical research aims at detecting, diagnosing, treating and preventing diseases. In order to carry out a meaningful research on a specific issue, one must layout the study design, collect the relevant data, outline the models and methods to analyze these data and finally fit the models to data and perform the analysis. Starting right from the layout of the study design until the derivation of the conclusions, statistics plays an indispensable role in biomedical research. Appropriate and adequate statistical analysis of biomedical data is of utmost importance in order to make reliable decisions targeting human population. The involvement of Statistics in medical research is gaining momentum and it is becoming more and more difficult to publish medical research if it is not well supported by statistical analysis.

This study explores the usage of various statistical methods; statistical software; complexity of statistical analysis; and inappropriateness of statistical methods over a decade and its improvements in publications of bio medical research.

Motivation for the study

It has been observed that Statistics has been gaining more and more recognition as an important component of medical research. Bio- statistics gives meaningful inference of uncertain and haphazard data, but for that one should have the knowledge of that. Almost all research papers are using statistical methods and techniques. Although Statistics contributes a lot in medical research, there is still a need for the in-depth conceptual understanding of the research problem on hand and the extent of and nature of statistical analysis required to attain desired task of research so that possible errors arising out of lack of conceptual clarity regarding appropriate statistical techniques applicable for carrying out comprehensive statistical analysis of the research problem can be avoided. Some of such common errors committed in medical research studies are as follows.

Some of the errors noticed in analysis of descriptive statistics:-

- No mention of the level of measurement for each variable
- Used "Mean± SD" for ordinal data.
- Didn`t mention the "±" sign for descriptive statistics for normally distributed data..
- Calculated mean and SD for ordinal data.
- Did not mention the sample size calculation and sampling technique.
- Only calculated measure of location (i.e. only mean)
- Wrong usage of graphs.

Some of the errors noticed in analysis of Statistical methods (Inferential statistics

- Parametric test applying without check distribution assumption of data
- Confusion regarding Parametric and Non parametric test.
- Use of independent t-test for paired data.
- Use of independent t-test for more than two independent groups.
- Wrong used *t*-test for comparing survival times (some of which are censored)

- Only t-test defined, did not mention about the problem which it is applied (i.e. independent, paired).
- Use of application t –test for comparison of more than two independent groups(mean)
- Confusion regarding application of statistical method (i.e. chie-square test or fisher`s exact test, t-test or Mann-Whitney test)
- Use of chi-square instead of fisher test /Yate`s correction
- Coefficient of agreement not computed.(i.e. kappa, Cramer, ….etc)
- Use of correlation coefficient when numbers are very small or when plotting of the data shows a apparently nonlinear trend.
- Power of the study not mentioned.
- For ANOVA no mention of degrees of freedom.
- Applied ANOVA without considering the assumption of required to apply it.
- Before applying ANOVA directly applied Post –hoc test.
- Usage of ANOVA for comparison of more than two related group mean.
- Used linear regression for ordinal data
- Comparison of p-values.
- Exact p-values not mentioned.(i.e. N.S….)
- Mention only p- value but didn`t write the name of statistical method.
- Only mentioned the p- value without calculate confidence interval.

Some other Errors noticed:-

- Statistical method mentioned only but did not mention p- value.
- Statistical Methods mentioned but not apply.
- Didn`t mention the name and version of statistical software which used in analysis.
- Incorrect Name statistical test was given.
- In Methodology section not included statistical method.

Therefore, it is felt that a detailed analysis of such errors should be carried so that such analysis can be very helpful to the present and future generations of researchers of medical sciences.

Aims and objective:-

1) The primary objective of this study is to assess study design and statistical techniques used and find out the type of error in different area of Bio medical research in the past decade and the present era.

2) The secondary objective of this study is to analyse the change in the trend of the use of statistical method or tools and type of errors taking place in such uses in the bio medical research.

Methodology

The basic methodology adopted for this analytical study is survey method. The basic components of it are as follows:

Sampling frame: -

The sampling frame for this study consists of list of all the volumes of all indexed bio-medical journals subscribed by and available at the central library of our institute Surat Municipal Institute of Medical Education and Research (SMIMER).

Sampling Technique:-

Considering the complex nature of the study covering several major areas and sub-area of medical research standard random sampling methods are not found appropriate for selecting a truly representative sample. Hence, the sampling method used for serving this most important purpose is Purposive Sampling.

Data Collection:-

In order to deal with the objective of this study, primary data have been collected from indexed peer reviewed bio medical journals subscribed by the central library of our institute (SMIMER). We have selected only original articles from them. This study scrutinized only the available research papers of latest issues/ latest volume of bio-medical publications during 2004 and 2013.

List of the journal and issues /volume included in the study given in Annexture-1

All papers were scrutinized thoroughly. Categories of statistical methods given by Emerson and Colditz[5] were used to classify statistical methods, with certain modification. Details of different categories are given in annexture-2.

As there were no statistical methods like cost benefit analysis, life table, etc these categories were not considered for final categorization table. If more than one statistical method were used in a single paper all were considered separately. If the same statistical method applied more than once in the same paper, it was counted only once.

Other information recorded for the both years (2004 and 2013) from these published research articles were: study design, appropriateness of study design, appropriateness of statistical methods mentioned, complexity of data analysis and usage of software given by Wang and Zhang, Qamruz Zaman et al[7], Thomas A. Langa and Douglas G. Altmanb[5] were used but with certain modifications.

Main outcomes of the study:-

Original published research articles were scrutinized from peer-reviewed journals of bio medical research in 2004 and 2013.It has been found that the prevalence of correctly mentioned study design is high in 2013 as compare to 2004, but it has been also observed that in the current era in very few original articles had mentioned the study design.

Sampling techniques are also mentioned in very few articles in both years but, the prevalence of mentioned sampling technique correctly is high in 2013 as compare to 2004.Commonly used probability sampling methods in medical researchers are simple random sampling, systematic sampling, stratified sampling and cluster sampling.

Bio-medical research is mainly related with human beings, so its very important to decide sample size while you plan the study. In this study ,it has been observed that more articles which published

in 2013 define the sample size as compared to 2004, but in very few articles it has been observes that the sample size calculation justification was given for both year (i.e for 2004 and 2013), authors mainly define the time period as sample size justification

This study covered total more than 1000 original published research articles of different area of biomedical research in a year 2004 and 2013, the usage of statistical methods (classification of statistical methods given in annexture-2) in research publications were significantly more in 2013 than 2004.

Usage of advance statistical method used more than one or two statistical methods and statistical software were increased over a decade. Chi square and independent t-test was used in both years commonly, but regression analysis (logistic) and survival analysis became more usable in 2013.

Many of researchers used concept of mix mathematical model, Yate`s correction, coefficient of agreement, Bootstrap in 2013 and it was the good improvement.

It has been also found that many of research articles used only descriptive statistics though they can use statistical methods.

Concerning complexity of statistical analyses (classification given in annexture-3) in 2004 more researchers used only basic statistical methods where as more advance techniques used in 2013.

It has been also observes that more number of articles used more than two or three statistical techniques in 2013 as compared to 2004.

Significant improvement was found in appropriate usage of study design and statistical methods in 2013 as compared to 2004, but it has been observed some common error in both years.

Limitations of the study:-

The review has been limited to statistical part of the manuscript only. Original research articles published in above listed journal were included in the study. As very few statistics were used in editorials, letters, case reports, short communications and other type of article, they were excluded.

References:-

1) Altman DG: Statistical reviewing for medical journals. Statist Med 1998;17:2661-2674.

2) Altman DG, Goodman S: Transfer of technology from statistical journals to the biomedical literature: past rends and future predictions. J AmMed Assoc 1994; 272:129-132.

3) Bland M. (1987) An Introduction to Medical Statistics OxfordUniversity Press, Oxford. (Reprinted 1988, 1989, 1990, 1991,1992, 1994, English Language Book Service edition for developing countries 1989.)

4) Boheng Zhang, MS; Qian Wang, MS , Research Design and Statistical Methods in Chinese Medical Journals, j*AMA*. 1998;280(3):283-285

5) John D. Emerson and Graham A. Colditz, Use of Statistical Analysis in The New England Journal of Medicine, New England Journal of Medicene 1983; 309:709-713

6) Lang TA, Altman DG. Basic Statistical Reporting for Articles Published in Biomedical Journals: The "Statistical Analyses andMethods in the Published Literature" or The SAMPL Guidelines" in: Smart P, Maisonneuve H, Polderman A (eds). Science Editors' Handbook, European Association of Science Editors, 2013

7) Masood Hussain Rao, Nazeer Khan,Comparison of statistical methods, type of articles and study design used in selected Pakistani medical journals in 1998 and 2007, J Pak Med Assoc, Vol. 60, No. 9, September 2010

8) Qamruz Zaman et al. Statistical methods and complexity of data analysis in recent surgical research Elixir Hum. Physio. 35 (2011)2961-2963.

9)Reed JF 3rd, Salen P, Bagher P.Methodological and statistical techniques: what do residents really need to know about statistics? Journal of Med syst ,2003 Jun;27(3):233-8.

10)Swati Patel, Prakash Patel, RK Bansal, Some Basic Application Of Bio-Statistics In Biomedical Research ,National Journal of Community Medicine,

11)Swati Patel, V D Naik, Prakash Patel, Use of statistical method and complexity of data analysis in recent research publications in basic medical sciences, National Journal of Community Medicine, Volume 5,Issue 2,Apr – June 2014

Proposed content of the thesis

This thesis will consist of the following chapters:

1. Introduction
2. Review of literature
3. Objective of study
4. Methodology
5. Results

5.1 Analysis of usage of medical research methodology (study design, sample size and sampling techniques) in basic and paramedical science and clinical (surgical and non surgical).

5.2 Analysis of usage of statistical methods and software, their inappropriateness, and their complexity of statistical methods in Basic Medical Science and paramedical science.

5.3 Analysis of usage of statistical methods and software, their inappropriateness, and their complexity of statistical methods, in Clinical (surgical and non surgical) Medical Science.

6. Discussion
7. Summary and conclusion.
8. Suggestions and Recommendations.
9. References

Annexures

Signature of Candidate

Signature of Supervising Teacher/Guide

(Swati .B. Patel)

(Dr.Vipin.D.Naik)
Principal of S.P.B collage of Commerce, Surat

Annexure -1

No.	Name of journal (2004-2013)
1.	Journal of Cytology, volume -30
2.	Indian Journal of Physiology and Pharmacology volume 48 and 57.
3.	Journal of applied Physiology ,volume 97 and 114
4.	Indian Journal of Clinical Biochemistry ,volume 19 and 28
5.	Indian Journal of Pathology and Microbiology ,Volume 56
6.	Indian Journal of Medical Microbiology, Volume 22 and 31
7.	Indian journal of Pharmacology Volume 36 and 45
8	Journal of Pharmacology and Pharmacotherapeutic Volume 4
10.	Indian Journal of Public Health, Volume 48 and 57
11.	Indian Journal of Community Medicine, Vol. 38, XXI
12.	The New England journal of Medicine Vol .351 and 369
13.	Indian Journal Critical Care Med January-March 2004 Vol 8 ,17
14	JAPI vol.52 ,61
15	Indian J Medical Research 119 ,137
16	Indian Journal of Nephrology vol 14,23
17	Indian Pediatric Vol 41 ,50
18	Indian Journal of Medical and Pediatric Oncology Vol 25 and 34
19	Journal of Indian Association of pediatric surgeons vol 9 and 18
20	Journal of Minimal Access Surgery Vol. 9 started by 2005
21	Indian journal of surgery Vol .75
22	Indian Journal of Urology vol 20 and 29
23	Clinical and Experimental Otorhinolaryngology vol. 6
24	Indian Journal of Orthopaedics vol. 38 and 47
25	Indian Journal of Psychiatry vol 46 and 55
26	Indian Journal of Radiology and Imaging vol 14 and 23
27	Indian Journal of Dermatology, Venereology, and Leprology vol 70 and 79
28	Indian Journal of Dermatology vol 58
29	Lung Indian vol 21 and 30
30	The Indian Journal of Chest Diseases & Allied Sciences Vol 46 and 55
31	Indian journal of tubercolosis vol 51 and 60
32	The Indian Journal of Chest Diseases & Allied Sciences Vol 46 and 57
33	Journal of Anesthesiology Clinical Pharmacology , Vol 29
34	Indian journal of Anesthesia ,Vol 48 and 56
35	Infect Disease of Obstet Gynecol 2013 ,vol .2003 &2013
36	The journal of obstetrics and gynecology of India vol.63.

Annexure:-2

- Sample size mentioned or not
- Study design mentioned or not
- Sampling technique mentioned or not
- Randomization mentioned or not
- Use descriptive or no statistical method (Mean, SD, Median, IQR, Mode,C0efficient of variation ect.)
- Check list of used statistical methods
  - t-test (one sample, independent and paired)
  - contingency table (Chi-square , Fisher exact test, McNemar's test)
  - Pearson`s correlation
  - Linear regression (least square method in which one predictor and other respondent)
  - Logistic regression
  - Analysis of Variance(ANOVA,ANCOVA, Repeated measure of ANOVA)
  - Multiple comparison (Procedures for handling multiple inferences on same data sets (e.g., Bonferroni techniques,Scheffe's contrasts, Duncan's multiple range procedures, Newmann-Keuls procedure)
  - Multiple regression (linear ,logistic)
  - Nonparametric test (Sign test, Wilcoxon signed ranks test, Mann- Whitney test, Kendall's tau, chi –square test for trend)
  - Nonparametric correlation (Spearman`s correlation)
  - Survival Analysis. (Kaplan-Meier survival curve, Cox regression, log rank, Cox model for comparing survival)
  - Epidemiological measures
  - (Relative risk, odds ratio, log odds, measures of association, sensitivity, specificity)
  - Normality
  - Others(ROC , Others regression analysis, principal component, Cluster analysis, Mathematical models)
  - 95% confidence interval
  - Specify one tail or two tail hypothesis
  - software-used for statistical methods

Annexure:-3

Basic statistics included: - t-test (one sample, independent t-test, paired t-test), contingency table (chi-square, fisher), correlation (Pearson, spearman),Nonparametric test

Modern statistics included:-contingency table (kappa), regression (linear, logistic), multiple regression, ANOVA, multiple comparison test (post hoc), Epidemiological analysis, Confidence interval,epidemiological analysis, Confidence interval